Wie exakt muss eine Wiedergabe erfolgen, um plausibel in einer virtuellen Umgebung zu wirken?

Robert Rosenkranz[1], M. Ercan Altinsoy[2], Kathleen Yen Li Kang[3], Shu-Chen Li[4]

[1,2]*Professur Akustik und Haptik, Technische Universität Dresden, 01062 Dresden,* [1] *E-Mail:robert.rosenkranz@tu-dresden.de*

[3,4]*Professur für Entwicklungspsychologie und Neurowissenschaft der Lebensspanne*

[1,2,4]*Centre for Tactile Internet with Human-in-the-Loop (CeTI)*

Einleitung

Ein wesentlicher Motivator für die Psychoakustik ist die Frage: „Wie kann Qualität vorhergesagt werden?“ [1] schlägt vor, das Qualitätsurteil als ein Maß für die Übereinstimmung zwischen Wahrnehmungsobjekt und Erwartungen bezüglich aller relevanter Eigenschaften in einem Anwendungskontext zu definieren. In virtuellen Umgebungen ist das Ziel, dass die die dargestellte Szene möglichst glaubwürdig bzw. plausibel wahrgenommen wird um als realistisch Empfunden zu werden. Dies ist natürlich ebenfalls von den Erfahrungen bzw. Erwartung abhängig. Das Plausibilitätsurteil lässt sich also ebenso definieren [2]. Man könnte also das Qualitätsurteil sowie das Plausibilitätsurteil bspw. als euklidischen Abstand d zwischen einem Wahrnehmungsobjekt s, welches durch einen Stimulus hervorgerufen wird, und den Erwartungen e formalisieren (siehe Abbildung 1). Alle relevanten Eigenschaften spannen dabei einen n-dimensionalen Wahrnehmungsraum auf, wobei die Wahrnehmungsmerkmale, a_1 bis a_n dessen Dimensionen repräsentieren. Damit müsste schließlich Qualität bzw. Plausibilität negativ mit diesem Abstand korreliert sein, also bspw. eine inverse Proportionalität annehmen. Daraus hat sich die Frage ergeben: Ist die Distanz im Wahrnehmungsraum ein starker Prädiktor für die Plausibilität? Welcher Zusammenhang besteht zwischen der Distanz und dem Plausibilitätsurteil?

Forschungskonzept

Aufbauend auf vorherigen Studien zur taktilen Wahrnehmung [3] insbesondere bezüglich Ganzkörperschwingungen im Fahrzeugkontext wurde daher folgendes Forschungskonzept zur Untersuchung der genannten Fragestellung überlegt (siehe Abbildung 1). Zunächst wurden einige Fahrzeugszenen ausgewählt, von den angenommen werden konnte, dass diese mit unterschiedlichen Erwartungen verknüpft sind. Für jede dieser Szenen wurden Vibrations-, Video und Audioaufnahmen durchgeführt. Diese Szenen wurden im Folgenden für zwei Untersuchungen verwendet. Zunächst wurden verbale Beschreibungen der Szenen definiert, bspw. „Stellen Sie sich vor sie würden mit einem Mittelklassefahrzeug eine Kopfsteinpflasterstraße mir 30 km/h befahren.“ Mit Hilfe dieser verbalen Beschreibungen wurde dann in

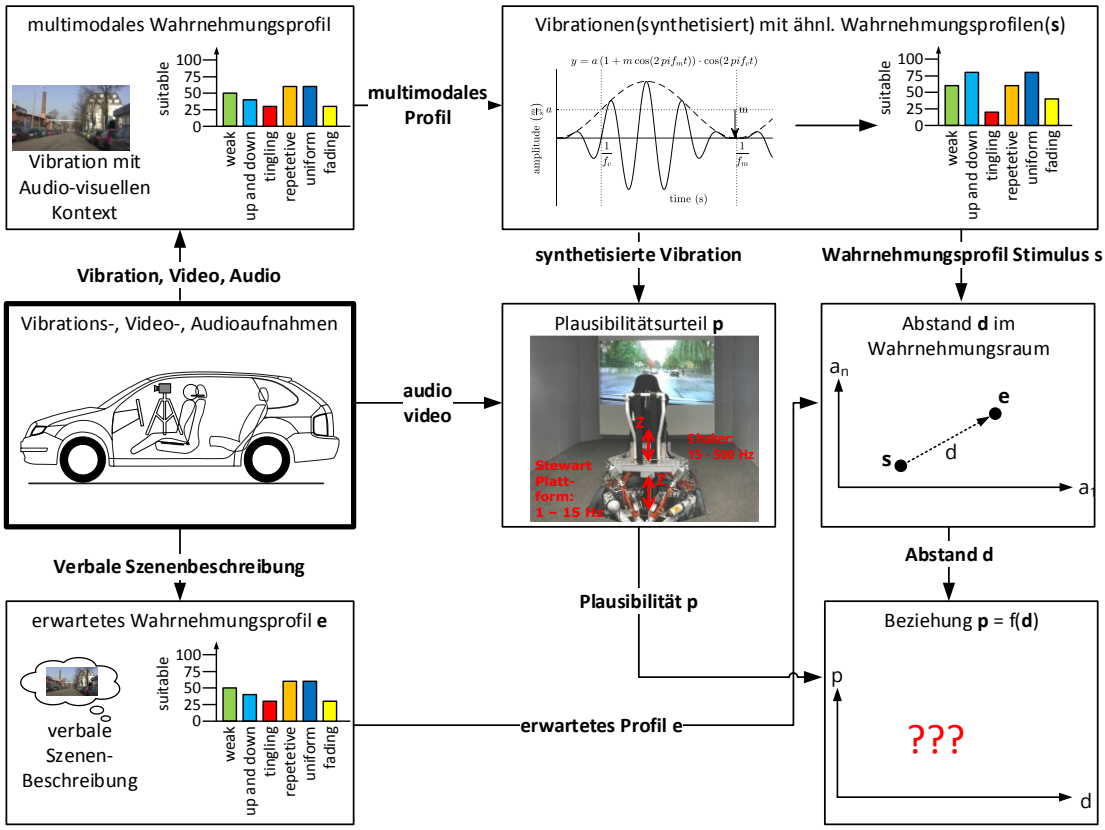

Abbildung 1: Forschungskonzept zur Untersuchung der Beziehung zwischen Plausibilität und der Distanz zwischen erwartetem Wahrnehmungsprofil und hervorgerufenem Wahrnehmungsprofil im sensorisch taktilen Wahrnehmungsraum

einer Befragung die erwarteten taktilen Wahrnehmungsprofile bestimmt. Diese bestehen aus den Wahrnehmungsmerkmalen schwach, auf und ab, kribbelnd, wiederholend, gleichmäßig und nachschwingend, welche die Dimensionen des sensorisch taktilen Wahrnehmungsraum hinreichend repräsentieren [3]. Parallel dazu wurden die aufgenommenen multimodalen Szenen in einer virtuellen Umgebung dargeboten und so das hervorgerufene, taktile Wahrnehmungsprofil bestimmt. Um die Beziehung zwischen der Distanz zu den Erwartungen im Wahrnehmungsraum und der Plausibilität zu untersuchen, waren Vibrationen mit unterschiedlichen Wahrnehmungsprofile notwendig. Dazu wurde auf die Datenbank mit Vibrations-Wahrnehmungsprofil-Paaren aus [3] zurückgegriffen. Darauf basierend wurde schließlich eine Wahrnehmungsuntersuchung durchgeführt, in der Vibrationen aus der Datenbank zusammen mit den Audio- und Videoaufnahmen der Fahrzeugszenen wiederum in einer virtuellen Umgebung präsentiert wurden. Es wurde jeweils die wahrgenommene Plausibilität der Vibrationen im Kontext der auditiv-visuellen Szene abgefragt. Die Distanz im Wahrnehmungsraum konnte nun als euklidischer Abstand zwischen erwartetem Wahrnehmungsprofil und den bereits aus der Datenbank bekannten Wahrnehmungsprofilen berechnet werden. Schließlich konnte die Beziehung zu den Plausibilitätsurteilen in Beziehung zur Distanz im Wahrnehmungsraum gesetzt werden.

Bestimmung des erwarteten und des hervorgerufenen Wahrnehmungsprofils

Um die eigentliche Forschungsfrage beantworten zu können, mussten erst einmal für unterschiedliche audiovisuelle Kontexte die erwarteten und die hervorgerufenen taktilen Wahrnehmungsprofile bestimmt werden.

Design

Dazu wurde jedes der sechs Wahrnehmungsmerkmale („auf und ab“, „kribbelnd“, „schwach“, „wiederholend“, „gleichmäßig“ und „nachschwingend“) aus [3] für alle 4 Fahrzeugszenen bewertet. Diese wurden jeweils durch verbale Beschreibung (Modus „erwartet“) und multimodale Darbietung (Modus „multimodal“) präsentiert. Um Priming durch die multimodale Darbietung zu vermeiden, wurden die Szenen im erwarteten Modus vor den Szenen im multimodalen Modus präsentiert. Unabhängig vom Präsentationsmodus wurde jedes Attribut für jede der 4 Szenen auf der gleichen quasi-kontinuierlichen Skala mit äquidistanten verbalen Ankern "nicht", "wenig", "mittel", "ziemlich" und "sehr" bei 0, 25, 50, 75 und 100 bewertet [4], die in einer MATLAB-GUI implementiert wurde. Alle Stimuli wurden in zufälliger Reihenfolge pro Modus präsentiert. Die Aufgabe im multimodalen Modus bestand darin, die tatsächlich wahrgenommenen Eigenschaften der Vibration der Szene zu bewerten. Die Aufgabe des verbalen Modus bestand darin, sich die erwarteten Eigenschaften der Vibration anhand der verbalen Beschreibung der Szene vorzustellen, z. B. "Stellen Sie sich vor, Sie fahren mit einem Mittelklassewagen über eine raue Kopfsteinpflasterstraße mit einer konstanten Geschwindigkeit von 30 km/h. Wie würde sich die Vibration anfühlen?".

Tabelle 1: Übersicht über die aufgezeichneten multimodalen Szenen

Nr.	Kategorie	v	Belag
3	nicht-transient	50	Kopfstein (fein)
6	nicht-transient	50	Kopfstein (grob)
7	nicht-transient	50	Asphalt (Landstraße)
11	nicht-transient	50	Asphalt

Stimuli

Die Stimuli bestanden aus Szenen mit einer Dauer von 4 s in einem Mittelklassewagen (Renault Scenic 1.6), die alltägliche Erfahrungen mit Ganzkörpervibrationen repräsentieren. Der Parameter Geschwindigkeit wurde auf 50 km/h festgelegt sowie der Betriebszustand auf Konstantfahrt, während die Straßenoberfläche systematisch variiert wurde, um häufige Situationen im Alltag zu repräsentieren. Eine Übersicht über die aufgezeichneten Szenen ist in Tabelle 1 zu finden. Zur Aufzeichnung der in den Fahrer eingeleiteten vertikalen Ganzkörperschwingungen wurden ein Sitzkissen-Beschleunigungssensor (B&K 4515B) und ein Kistler-Beschleunigungssensor 8305B10 verwendet. Für die Videoaufnahmen wurde eine Canon EOS 600D Kamera mit optischem Bildstabilisator verwendet. Der Ton wurde mit zwei B&K 2671 ICP-Mikrofonen aufgenommen, die an der Kopfstütze neben den Ohren des Fahrers angebracht waren. Die multimodalen Szenen setzten sich aus diesen Aufnahmen zusammen. Die Stimuli des erwarteten Modus bestanden aus verbalen Beschreibungen des Szeneninhalts: Fahrzeugtyp, Geschwindigkeit, Betriebszustand (konstante Geschwindigkeit), Straßenoberfläche.

Versuchsaufbau

Das Experiment wurde in einem Virtual-Reality-Labor [5] durchgeführt, in dem Bild, Ton und Vibration wiedergeben werden können (Abbildung 2). Die aufgezeichneten vertikalen Ganzkörperschwingungen wurden mit einer pegelabhängigen Frequenz von 7 bis 15 Hz getrennt. Die tiefpassgefilterten Schwingungen wurden mit einer Steward-Motion-Plattform und die hochpassgefilterten Schwingungen mit einem elektrodynamischen Shaker präsentiert. Die potenziellen individuellen Unterschiede in der Vibration, die verschiedenen Versuchspersonen z. B. aufgrund ihres Gewichts dargeboten wurden, wurden mit einem FIR-Filter [6] kompensiert, um die aufgezeichnete Vibration genau zu reproduzieren. Ein Projektor wurde verwendet, um die Videos der Szene für den multimodalen Modus zu präsentieren. Der Ton wurde im multimodalen Modus mit einem Wellenfeldsynthesesystem mit 464 Lautsprechern als fokussierte Schallquellen an den Ohren der Probanden präsentiert.

Probanden

Insgesamt 22 Probanden (13 Männer, 9 Frauen) mit einem Durchschnittsalter von 30 Jahren (19 bis 61 Jahre) nahmen an dem Experiment teil. Die Studie wurde mit dem Einverständnis und der schriftlichen Zustimmung aller Teilnehmer durchgeführt.

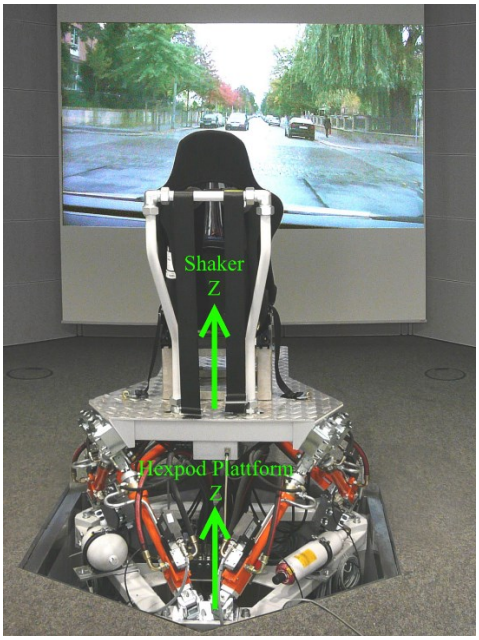

Abbildung 2: Wiedergabesystem zur Erzeugung von auditiv-visuell-taktilen Stimuli

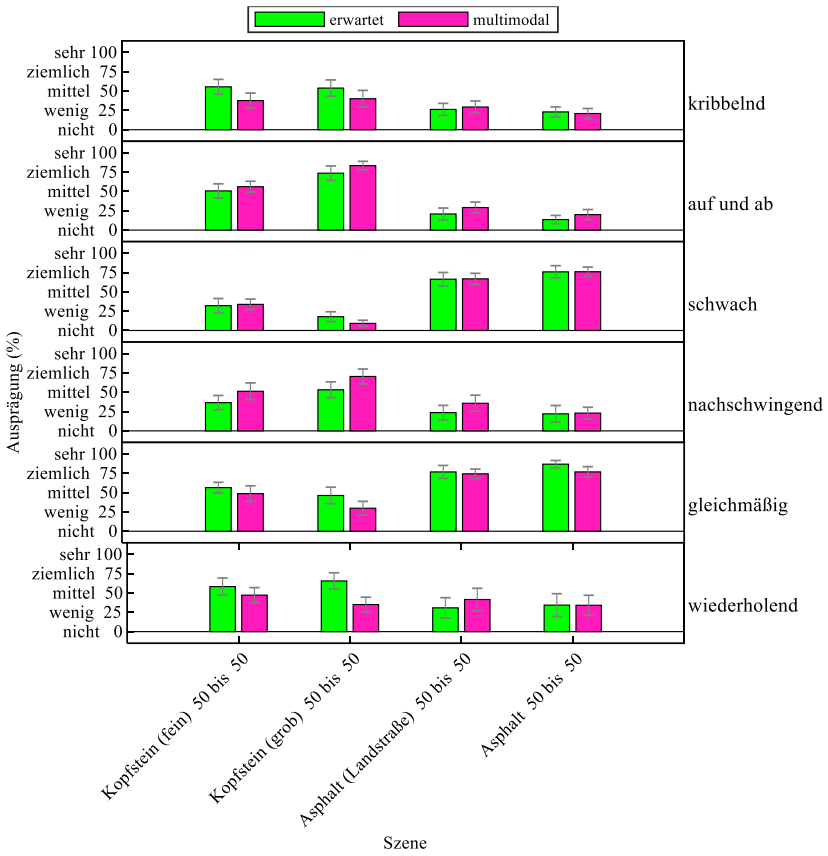

Abbildung 3: Mittelwerte und 95%-Konfidenzintervalle der Wahrnehmungsmerkmale für jede Szene und beide Präsentationsmodi.

Ergebnisse

Abbildung 3 zeigt ein Diagramm mit den mittleren Bewertungen für jedes Wahrnehmungsmerkmal und den 95%-Konfidenzintervallen alle 4 Szenen. Der mittlere Bewertungsunterschied zwischen dem erwarteten Modus und dem multimodalen Modus über alle Szenen und Wahrnehmungsmerkmale liegt bei nur 2 Punkten auf der 100-Punkte-Skala. Der Bereich der Konfidenzintervalle für den erwarteten Modus liegt bei 4 bis 16 Skalenpunkten und für den multimodalen Modus bei 4 bis 15 Punkten. Ein paarweiser Kontrast für den Unterschied zwischen multimodalem und erwartetem Modus ergab ein 95%-Konfidenzintervall für die Effektgröße von -1 bis 4 Punkten. Es ist also unwahrscheinlich, dass der kleine Unterschied zwischen den Präsentationsmodi ein Fünftel des 25-Punkte-Skalenintervalls oder 1/25 der 100-Punkte-Bewertungsskala überschreitet. Dies deutet darauf hin, dass die erwarteten Wahrnehmungsprofile durchaus repräsentativ für die multimodalen Szenen sind.

Beziehung zwischen der Plausibilität und der Distanz zwischen erwarteten und hervorgerufenem Wahrnehmungsprofil

Design, Stimuli und Versuchsaufbau

Um nicht aufgezeichnete Vibrationen und synthetischen Vibrationen zu mischen, wurden nun aus der Datenbank [3] zunächst die Vibrationen mit dem Wahrnehmungsprofil ausgewählt, dessen euklidischer Abstand minimal zu dem durch die aufgenommen Szenen hervorgerufenen Wahrnehmungsprofil ist. Die synthetischen Vibrationen wurden nun einfach in 9 Stufen im Vibrationspegel variiert. Dabei wurde abhängig vom Ausgangspegel die Vibrationen entweder abgeschwächt oder verstärkt. Für jede synthetische Vibration war dabei jeweils das hervorgerufenen Wahrnehmungsprofil aus der Datenbank bekannt. Mit dem im vorherigen Abschnitt verwendeten Versuchsaufbau wurden die ausschließlich die synthetisierten Vibrationen jeweils auf einer 100-Punkte-Skala nach ihrer wahrgenommenen Plausibilität im Kontext der auditiv-visuellen Szene bewertet. Die Plausibilität aller Szenenkombinationen wurde auf der im vorherigen Versuch beschriebenen Skala bewertet.

Probanden

Insgesamt 46 Probanden (27 Männer, 19 Frauen) mit einem Durchschnittsalter von 24 Jahren nahmen an dem Experiment teil. Die Studie wurde mit dem Einverständnis und der schriftlichen Zustimmung aller Teilnehmer durchgeführt.

Ergebnisse

In Abbildung 4 sind jeweils die Mittelwerte der Plausibilitätsurteile sowie die 95% Konfidenzintervalle dargestellt. Man sieht deutlich, dass das den Erwartungen ähnlichste Wahrnehmungsprofil auch als am Plausibelsten bewertet wird. In Abbildung 5 sind schließlich die Plausibilitätsurteile in Beziehung zur euklidischen Distanz im sensorisch taktilen Wahrnehmungsraum zwischen dem erwarteten Wahrnehmungsprofil und dem hervorgerufenen Wahrnehmungsprofil gesetzt. Man kann eindeutig eine inverse Korrelation zwischen dem Plausibilitätsurteil und deiser Distanz im Wahrnehmungsraum sehen.

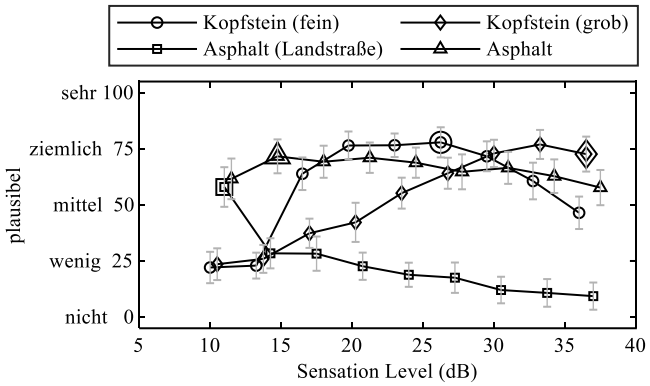

Abbildung 4: Mittelwerte und 95%-Konfidenzintervalle der Plausibilität der Stimuli in Abhängigkeit des fühlschwellenbezogenen Pegels

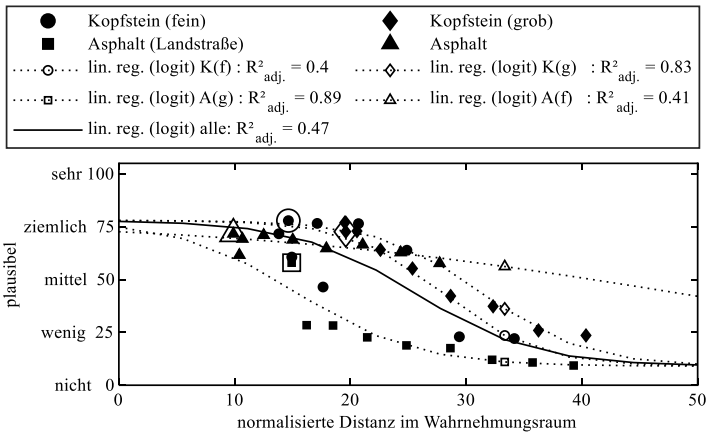

Abbildung 5: Mittelwerte der Plausibilitätsurteile in Abhängigkeit der Distanz zwischen erwartetem Wahrnehmungsprofil und hervorgerufenem Wahrnehmungsprofil.

Das bedeutet, dass je mehr die hervorgerufenen Wahrnehmungsprofile den erwarteten Wahrnehmungsprofilen ähneln, desto Plausibler werden die Vibrationen empfunden, welche diese Wahrnehmungsprofile hervorrufen. Man sieht außerdem, dass mit steigender Distanz die Plausibilität zunächst hoch bliebt, dann ab ca. einer Distanz von 15 steil abfällt und dann nahe am Minimum verbleibt. Allerdings scheint die Plausibilität bei einigen Szenen schneller abzufallen als bei anderen. Zur quantitativen Analyse der Beziehung wurde eine lineare Regression durchgeführt. Es ist zu erwarten, dass die Plausibilität nicht größer ausfallen kann, wenn die Distanz im Wahrnehmungsraum bereits nahe 0 ist. Ebenso konvergiert natürlich die Plausibilität mit steigender Distanz anscheinend gegen 0. Um diese Skalennutzungseffekte zu berücksichtigen, wurden daher die Plausibilitätsurteile zunächst Logit transformiert. Über alle Szenen hinweg beschreibt eine einfache lineare Regression die Beziehung mit einem R^2 von rund 0.5, wobei der Zusammenhang für eine Szene sogar rund 0,9 beträgt. Sofern auch Vibrationen mit größere Distanzen eingeflossen wären, würde das Bestimmtheitsmaß, welches ja die Verbesserung der Modellvorhersage gegenüber dem Mittelwert beschreibt, wahrscheinlich sogar noch größer ausfallen, z.B. bei der Szene mit feinem Asphalt.

Diskussion

Die experimentellen Ergebnisse bestätigen prinzipiell die Formalisierung der theoretischen Plausibilitäts- bzw. Qualitätsdefinition als Distanz im sensorisch-taktilen Wahrnehmungsraum zwischen erwarteten Wahrnehmungsprofil und durch Vibration hervorgerufenem Wahrnehmungsprofil. Eine lineare Regression kann jedoch noch nicht die gesamte beobachte Varianz erklären. Weitere Untersuchungen sollten also bspw. mit Variation anderer Vibrationsparameter zur Verbesserung des Modells beitragen. Weitere auditiv-visuelle Kontexte wären zur Klärung möglicher kontextabhängiger Parameter notwendig. Diese Formalisierung von Qualität und Plausibilität ermöglicht prinzipiell Vorhersagen, jedoch sind zusätzlich Modelle der einzelnen sensorisch taktilen Wahrnehmungsmerkmale dafür erforderlich.

Danksagung

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der Länder – EXC 2050/1 – Projektnummer 390696704 – als Exzellenzcluster „Centre for Tactile Internet with Human-in-the-Loop“ (CeTI) der Technischen Universität Dresden.

Literatur

[1] U. Jekosch, “Basic Concepts and Terms of ‘Quality’, Reconsidered in the Context of Product-Sound Quality,” *acta acustica united with Acustica*, vol. 90, no. 6, pp. 999–1006, 2004.

[2] K. Kuhn-Rahloff, “Prozesse der Plausibilitätsbeurteilung am Beispiel ausgewählter elektroakustischer Wiedergabesituationen,” Technische Universität Berlin, 2011.

[3] R. Rosenkranz and M. E. Altinsoy, “Mapping the Sensory-Perceptual Space of Vibration for User-Centered Intuitive Tactile Design,” *IEEE Transactions on Haptics*, vol. 14, no. 1, pp. 95–108, 2021.

[4] B. Rohrmann, “Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung,” *Zeitschrift für Sozialpsychologie*, vol. 107, no. 5, pp. 222–245, 1978.

[5] M. E. Altinsoy, U. Jekosch, J. Landgraf, and S. Merchel, “Progress in Auditory Perception Research Laboratories - Multimodal Measurement Laboratory of Dresden University of Technology,” in *Audio Engineering Society Convention 129*, 2010, pp. 1–8.

[6] M. E. Altinsoy and S. Merchel, “BRTF (Body-Related Transfer Function) and Whole-Body Vibration Reproduction Systems,” in *Audio Engineering Society Convention 130*, 2011, pp. 1–5.